Optional Homework Problems

The attached page 445 is from the Hays (1973) statistics book. The page contains formulas for the density of F, and the mean and the variance of F.

1. Write a SAS or SPSS program to compute the mean and variance of F for any value of numerator, ν_1, or denominator, ν_2, degrees of freedom.

2. Write a program for the density function of the random variable F. I am not sure whether SAS or SPSS has a factorial function so check for that first. Note that value of the density function is for a given value of F and numerator, ν_1 or denominator, ν_2, degrees of freedom. The density function will give the height of the F distribution for a given value of F.

3. SAS (Statistical Analysis System) has functions to find an F value given the probability "p", and numerator "ndf" and denominator degrees of freedom "ddf" and the noncentrality parameter "nc"and a function to find the probability of a certain F with numerator degrees of freedom, denominator degrees of freedom, and the noncentrality parameter. The commands are summarized below:

SAS command to find F for a certain probability.
F=FINV(p,ndf,ddf,nc);

SAS command to find the probability of an F value.
P=PROBF(F,ndf,ddf,nc);

a. Write a program to compute the probability for any value of F.

b. Write a program to compute the F for any p, ndf,ddf, and nc.

c. What does the program below do? 210 is the SS_A and 14.17 is the MS_A for the Keppel one factor ANOVA example that we have used in class.

```
data retpower;
noncent=210/14.17;
fcrit=finv(.95,2,12,0);
power=1-probf(fcrit,2,12,noncent);
proc print;
run;
```

have the same variance, and that the samples drawn are independent, then the theoretical distribution of F values can be found.

Formally, the density function for the random variable F is defined by

$$f(F;\nu_1,\nu_2) = \begin{cases} \left[\dfrac{(\frac{1}{2}\nu_1 + \frac{1}{2}\nu_2 - 1)!}{(\frac{1}{2}\nu_1 - 1)!(\frac{1}{2}\nu_2 - 1)!}\right] \dfrac{F^{(1/2)\nu_1 - 1}(\nu_2/\nu_1)^{(1/2)\nu_2}}{(F + \nu_2/\nu_1)^{(1/2)\nu_1 + (1/2)\nu_2}} \\ \qquad \text{for } 0 < \nu_1,\ 0 < \nu_2,\ 0 \leq F, \\ 0 \qquad \text{otherwise.} \end{cases}$$

(Note that here, as in sections 10.4 and 11.2, we are violating our conventions about symbolizing random variables). Obviously, this function is much too complicated for us to gain an impression of the form of a given distribution from the function rule alone. However, it is worth noting that for the values of ν_1 and ν_2 that will usually concern us (that is, integral values, with $\nu_1 < \nu_2$) the F distribution is skewed positively, and is unimodal for $\nu_1 > 2$. On the other hand, the distribution can take on other forms when other conditions are set for ν_1 and ν_2.

The mean and the variance of an F distribution also depend on the values of ν_1 and ν_2. Thus,

$$E(F) = \frac{\nu_2}{\nu_2 - 2} \qquad \text{for } \nu_2 > 2$$

E(F)

and

$$\mathrm{Var}(F) = \frac{2\nu_2^2(\nu_1 + \nu_2 - 2)}{\nu_1(\nu_2 - 2)^2(\nu_2 - 4)} \qquad \text{for } \nu_2 > 4.$$

Var(F)

Note that as the value of ν_2 is made increasingly large, the mean of the F distribution approaches 1.00.

Mathematically, the F distribution *is* rather complicated. However, for our purposes it will suffice to remember that the density function for F depends only upon two parameters, ν_1 and ν_2, which can be thought of as *the degrees of freedom associated with the numerator and the denominator of the F ratio*. The range for F is *nonnegative real numbers*. The expectation of F is $\nu_2/(\nu_2 - 2)$ for $\nu_2 > 2$. In general form, the distribution for any fixed ν_1 and ν_2 is nonsymmetric, although the particular "shape" of the function curve varies considerably with changes in ν_1 and ν_2.

Before we can apply this theoretical distribution to an actual problem of comparing two variances, the use of F tables must be discussed.

11.12 The use of F tables

Since the distribution of F depends upon two parameters ν_1 and ν_2, it is even more difficult to present tables of F distributions than those of χ^2 or t. Tables of F are usually encountered only in drastically condensed form.

Such tables give only those values of F that cut off the *upper* proportion Q in an F distribution with ν_1 and ν_2 degrees of freedom. The only values of Q given are those that are commonly used as α in a test of significance (that is, the